

Structure-based experimental confirmation of biochemical function to a methyltransferase, MJ0882, from hyperthermophile *Methanococcus jannaschii*

Lan Huang^{†1}, Liwei Hung^{†2}, Mark Odell¹, Hisao Yokota³, Rosalind Kim³ and Sung-Hou Kim^{*3,4}

¹Dept. of Cellular Biochemistry and Biophysics, Sloan Kettering Institute, 1275 York Ave., New York, NY 10021; ²Current address: Brookhaven National Laboratory, Upton, New York, 11973; ³Physical Biosciences Division, Lawrence Berkeley National Laboratory; ⁴Department of Chemistry, University of California, Berkeley, CA 94720, USA; *Author for correspondence (e-mail: shkim@Cchem.Berkeley.edu; tel: 510-486-4333; fax: 510-486-6059)

[†]These authors contributed equally to the work

Received 8 November 2001; Accepted in final form 17 March 2002

Key words: *Methanococcus jannaschii*, methyltransferase, Rossmann fold, structural genomics, X-ray crystallography

Abstract

We have determined the three-dimensional (3-D) structure of protein MJ0882, which derives from a hypothetical open reading frame in the genome of the hyperthermophile *Methanococcus jannaschii*. The 3-D fold of MJ0882 at 1.8 Å highly resembles that of a methyltransferase, despite limited sequence similarity to any confirmed methyltransferase. The structure has an S-adenosylmethionine (AdoMet) binding pocket surrounded by motifs with similarities to those commonly found among AdoMet binding proteins. Preliminary biochemical experiments show that MJ0882 specifically binds to AdoMet, which is the essential co-factor for methyltransferases.

Introduction

Methyltransferase has an α/β type structure with a central mixed β -sheet with several α -helices arranged around it. The sheets are formed by five parallel β -sheets with strand order 54123, and one or two anti-parallel β -hairpins next to β_5 [1]. The order of the parallel strands is reversed once with a switch point between β_4 and β_1 . The 3-D structures of a few methyltransferases have been previously determined, such as several DNA methyltransferases, HhaI [2, 3], TaqI [4], HaeIII [5] and PvuII [6], two small molecule methyltransferases, COMT [7] and GNMT [8], two RNA methyltransferases, VP39 [9] and ErmC' [10], and one protein methyltransferase, CheR [11].

In this paper, we describe the structure of a hypothetical protein, MJ0882, from the hyperthermophile *Methanococcus jannaschii* (MJ), and the importance of the 3-D structure in inferring the molecular func-

tion of a protein with unknown function. MJ0882 is a member of a sequence family that shares limited sequence homology to any proteins with known functions. Recent BLAST search yielded similar protein sequences of putative RNA methyltransferases, UPF0020 family [12]. Upon comparing the structure of MJ0882 with those in the protein data bank (PDB) using the DALI method [13], we found that the three-dimensional structure of MJ0882 is highly similar to several methyltransferases, especially to catechol o-methyltransferase [7], despite the absence of any significant sequence similarities. Transfer of the methyl group from AdoMet is an alkylation reaction central to cellular biochemistry [1]. Preliminary biochemical experiments show that MJ0882, in fact, binds to AdoMet.

Methyltransferases are central for a number of important biological processes, such as cell cycle control, signaling, and genetic imprinting. Transferring

the methyl group from AdoMet to either nitrogen, oxygen or carbon atoms is frequently employed in diverse organisms ranging from bacteria to plants and mammals. The substrates of this large class of enzyme include small molecules, DNA, RNA, protein, lipids, and polysaccharides. The mis-regulation of methyltransferases occurs in a diverse set of diseases, such as cancer. Thus, methyltransferases are suitable for therapeutic intervention for disease development. Our discovery that MJ0882 represents a new class of methyltransferase will open doors to the functional assignment of a large group of unknown proteins with sequences homologous to that of MJ0882, such as the UPF0020 family.

Materials and methods

Expression, purification, and crystallization of the MJ0882 protein

The DNA coding region for MJ0882 was prepared by polymerase chain reaction (PCR), cloned into the pET-21a vector (Novagen), and transformed into the *Escherichia coli* host BL21(DE3)/p SJS1244 [16] for isopropyl- β -D-thiogalactoside-dependent protein expression. The yield of protein was approximately 60 mg/L of culture. Because MJ is a heat-tolerant thermophile, the expressed protein MJ0882 was purified initially by an 80 °C incubation for 25 min in 50 mmol/L MES (2-(N-morpholino)ethane sulfonic acid) buffer at pH 6.2, 2 mmol/L EDTA (ethylene diamine tetra-acetic acid), 2 mmol/L DTT (dithiothreitol), 10% glycerol, followed by centrifugation. The MJ0882 protein was further purified by cation exchange chromatography over a Source S column (Pharmacia) pre-equilibrated with 50 mmol/L MES buffer at pH 6.2, and eluted with a 0 to 1 mol/L NaCl linear salt gradient. The protein eluted at 0.7 mol/L NaCl. The crystallization condition was initially determined with the sparse matrix method [17] using the Hampton Research Crystal Screen (Laguna Niguel, CA). Crystals for structural analysis were grown by the vapor-diffusion method from a solution containing 50 mg/ml of protein, 50 mmol/L MES, pH 5.13, 1.375 mol/L ammonium sulfate in a drop equilibrated against 0.1 mol/L MES pH 5.13, 2.75 mol/L ammonium sulfate. Selenomethionine (Se-Met)-substituted protein was made in a methionine auxotroph (B834(DE3)/pSJS1244) grown in minimal media in the presence of Se-Met, and the crystals were grown

under identical conditions to that of the native protein. The crystals were transferred into cryo-solvent containing the above crystallization mother liquor plus 25% cryostoil, and flash-frozen in a cold nitrogen stream (Oxfordcryo) set at 100 K prior to exposure to X-rays. Both the native and the Se-Met substituted crystals are in space group P21212, with unit cell dimensions of $a = 69.34 \text{ \AA}$, $b = 36.78 \text{ \AA}$, and $c = 68.78 \text{ \AA}$ at 100 K. Assuming one molecule per asymmetric unit, the estimated solvent content is 34.5%.

Data collection, structure determination and refinement

A three-wavelength MAD [14] X-ray diffraction data set was collected at the Macromolecular Crystallography Facility at the Advanced Light Source of the E. O. Lawrence Berkeley National Laboratory. The data was processed with DENZO and reduced with SCALEPACK [18]. The heavy atom searching and phasing were conducted with the program SOLVE [19] (www.solve.lanl.gov). Subsequent density modification was carried out with the program DM [20]. The initial atomic model was built with the program O [21]. A starting model containing residues 5–197 was refined against maximum-likelihood target using CNS [22] with bulk solvent correction. Conventional positional refinement, torsion angle dynamics simulated annealing [22] and individual atomic temperature factor refinement were cycled through the courses of refinement. The current model contains residues 4–197 and 163 water molecules, with an R-factor of 18.8% and R-free of 23.5%. The crystallographic and refinement statistics are shown in Table 1. The atomic coordinates and structure factors have been deposited in the Brookhaven Protein Data Bank under the Accession code 1DUS.

Binding studies of MJ0882 with AdoMet, AdoHcy, and NAD using a Gel filtration assay

MJ0882 at a concentration of 21.5 $\mu\text{mol/L}$ was incubated with 1.8 $\mu\text{mol/L}$ ^3H -labeled AdoMet for 5 min at room temperature in buffer A (20 mmol/L Bis-Tris propane at pH 6.5, 10 mmol/L MgCl_2 , 5 mmol/L DTT) in a final reaction volume of 50 μl . The AdoMet concentration was increased to 3.7 $\mu\text{mol/L}$ by the addition of unlabeled AdoMet immediately prior to applying 20 μl of the reaction mixture on to 700 μl Sephadex G-50 columns (Pharmacia), which

Table 1. Statistics from the crystallographic analysis.

Space group:	P2 ₁ 2 ₁ 2		
Cell parameters:	69.340 Å, 36.780 Å, 68.340 Å, 90°, 90°, 90°		
Data set	MJ0882	MJ0882	MJ0882
	SeMet λ1	SeMet λ2	SeMet λ3
Wavelength (Å)	0.9800	0.9797	0.9686
Beamline	ALS (5.0.2)	ALS	ALS
Resolution (Å)	1.8	1.8	1.8
Data coverage (%)	99.9	99.9	99.9
Data redundancy	8.0	8.0	8.0
R _{sym} %	7.1	7.6	7.3

Refinement statistics:						RMSD		
Data resolution set (Å)	Reflections ($ F > 2\sigma$)	Total atoms	Water atoms	R-factor (%)	R-free (%)	Bonds (Å)	Angles (°)	B-factor (Å ²)
20.0–1.8	16370	1534	163	18.8	23.5	0.010	1.70	2.5

$R_{\text{sym}} = \sum_h \sum_i |I_{h,i} - \langle I_h \rangle| / \sum_h \sum_i I_{h,i}$ for the intensity (I) of i observations of reflection h . R-factor = $\sum |F_{\text{obs}} - F_{\text{calc}}| / \sum |F_{\text{obs}}|$, where F_{obs} and F_{calc} are the observed and calculated structure factors, respectively.

Figure of merit = $|F(hkl)_{\text{best}}| / |F(hkl)|$. R-free = R-factor calculated using 10% of the reflection data chosen randomly and omitted from the start of refinement. RMSD: rmsd from ideal geometry and in B-factor of bonded atoms.

had been pre-equilibrated with two column volumes of buffer A containing 3.7 $\mu\text{mol/L}$ unlabeled AdoMet. The column eluate was collected in 40 μl fractions, 20 μl aliquots of which were measured for ^3H radioactivity by scintillation counting. The fractions containing significant amounts of the putative methyltransferase were identified by SDS-PAGE. The gel filtration profile of AdoMet alone was determined by preincubating the AdoMet with buffer A for 5 min prior to application to an identical Sephadex G-50 column. The same experiment was repeated with a 100 : 1 excess of cold AdoHcy, or a 100 : 1 excess of cold NAD, and the gel filtration profile was determined as above.

Accession numbers

Atomic coordinates and structure factors have been deposited with the Protein Data Bank as entry 1DUS.

Results and discussion

Structural comparison

The crystal structure of MJ0882 was determined at 1.8 Å resolution using the MAD method [14]. The current refined model contains residues 4 to 197 and 163 water molecules (Table 1). The structure consist-

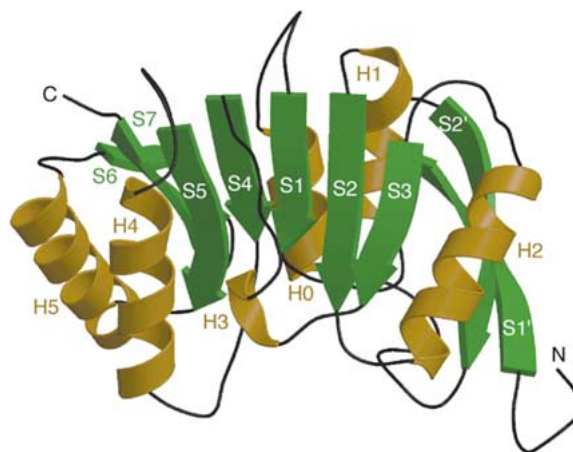


Figure 1. Three-dimensional structure of MJ0882 consists of 9 β strands and 5 α helices [prepared with the program MOLSCRIPT [23]]. The α -helices and β -strands are labeled and colored in yellow and green, respectively. N and C-termini are labeled.

ing of 8 β -strands and 5 α -helices resembles a Rossman fold (Figure 1).

To predict a possible molecular function for MJ0882, we compared its crystal structure with all representatives of protein folds in the PDB using the DALI method [13]. The proteins with DALI Z scores above 12 ($Z < 2$ usually means negligible similarity) include catechol o-methyltransferase (COMT) (PDB accession number 1vid, $Z = 14.0$), glycine n-methyltransferase (PDB accession number 1xva, $Z = 13.0$),

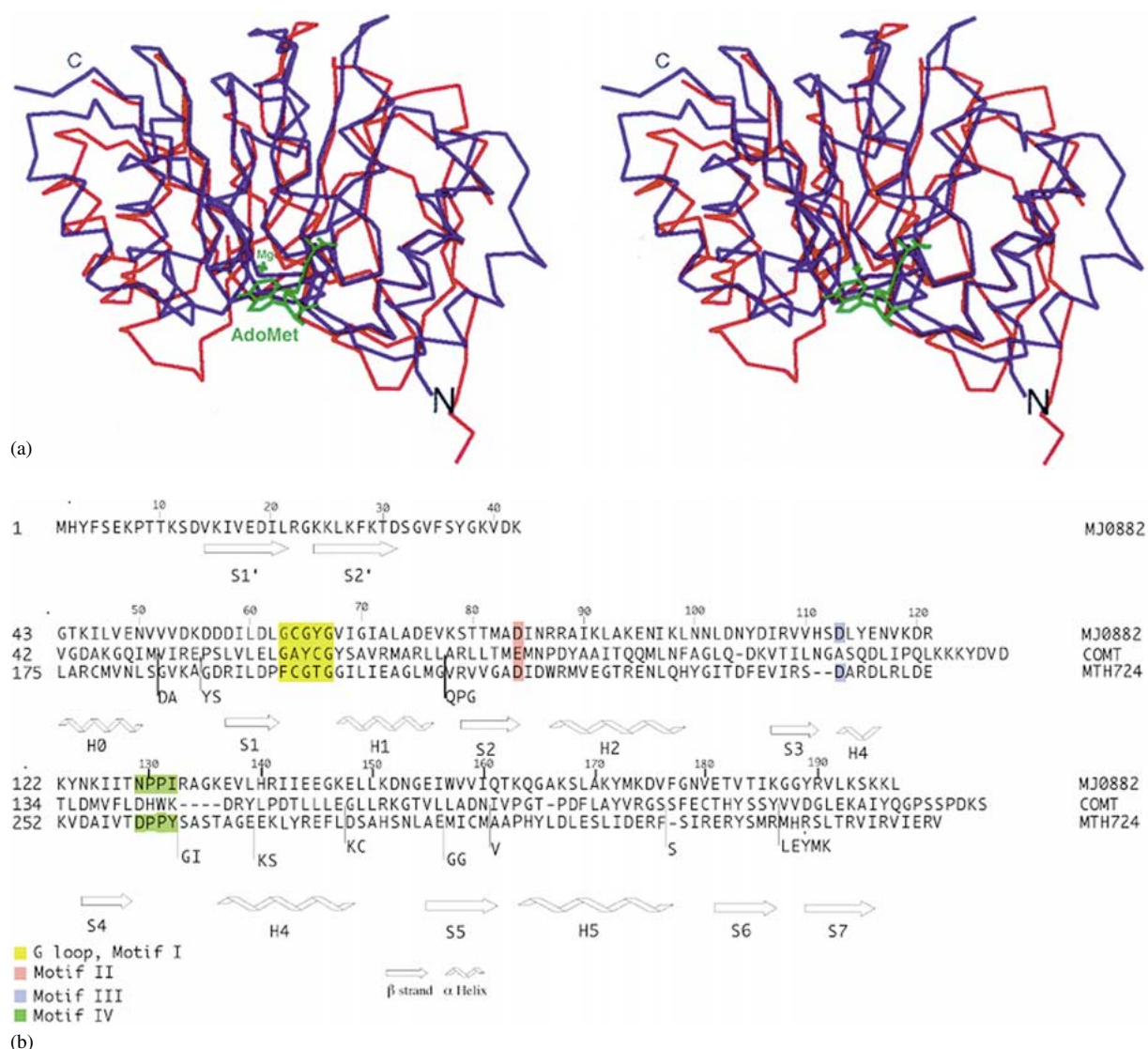


Figure 2. Structural and sequence alignment of MJ0882 and other methyltransferases. (a) Stereo view of the structural alignment of MJ0882 and COMT. MJ0882 and COMT are colored in red and blue, respectively. AdoMet and Mg^{2+} are colored green [prepared with the program INSIGHT II (Molecular Simulation Inc)]. (b) Sequence alignment of MJ0882, COMT, and putative RNA methyltransferase MTH724. Motifs important for AdoMet binding are colored and labeled.

adenine-n6-DNA-methyltransferase, TaqI (PDB 2adm, $Z = 12.9$), and chemotaxis receptor methyltransferase CheR (PDB 1af7, $Z = 12.1$). Cytosine-specific methyltransferase HhaI (PDB 6mht) scores at $Z = 9.6$. In addition, MJ0882 shows high structural similarity with NAD-dehydrogenase family. MJ0882 and COMT show high structural similarity (Figure 2a), despite the fact that the sequence identity between them is only about 5% (Figure 2b). The root-mean-square deviation (rmsd) of the $C\alpha$ positions of all the corresponding residues in the conserved seven-

stranded β -sheet region between MJ0882 and COMT (160 residues out of 188 residues of MJ0882) is 2.6 Å.

The MJ0882 structure has an AdoMet binding pocket surrounded by residues similar to those generally found among AdoMet binding proteins [1], as also in DNA methyltransferase TaqI. Methyltransferases possess four signature motifs [1]: (1) The 'G-loop' between $\beta 1$ and $\alpha 1$ which facilitates AdoMet binding; (2) the D-loop with an acidic residue (Asp or Glu) at the end of the $\beta 2$ whose side chain hydro-

gen-bonds with the ribose hydroxyl of AdoMet; (3) an Asp commonly located at the N-terminal residue of $\alpha 3$; and (4) the P-loop (proline rich) at the c-terminal of $\beta 4$ for DNA methyltransferases [15]. In MJ0882 the G-loop between $\beta 1$ and $\alpha 1$ corresponding to GCGYG (G-loop in COMT) engages in hydrophobic packing against the adenine ring in AdoMet. MJ0882 contains an acidic residue, Asp84, at the end of $\beta 2$ (D-loop). Asp 113 resides at the N-terminus of $\alpha 3$. The P-loop C-terminal of $\beta 4$ is NPPI, which is similar to the DNA methyltransferase M. TaqI (NPPY). The P-loop also participates in hydrophobic packing against the adenine ring of the AdoMet. The overall structure of MJ0882 is similar to that of the AdoMet binding domain of TaqI. The rmsd between MJ0882 and TaqI (143 residues out of 188 residues of MJ0882) is 2.9 Å. The AdoMet binding pocket of MJ0882 is also similar to that of TaqI (Figure 3). Ile 85 located between $\beta 2$ and $\alpha 2$ is similar to Ile72 in TaqI, which makes a face-to-face van der Waals contact with the adenine ring of AdoMet. Additionally, Tyr66 in MJ0882 forms an edge-to-face contact with the adenine ring of AdoMet, similar to Phe146 ($\alpha 4$) in TaqI DNA methyltransferase or Phe18 ($\beta 1$ - $\alpha 1$) in HhaI DNA methyltransferase. Thus, the structural comparison of MJ0882 with known methyltransferases suggests that MJ0882 may bind to AdoMet, the common methyl donor for methyltransferases. The following section presents preliminary biochemical evidence for the binding of AdoMet to MJ0882 *in vitro*. Proteins possessing Rossman folds are also known to bind to NAD or NADP as cofactors. Thus we designed additional experiments to show that MJ0882 specially binds to AdoMet, not to NAD or NADP.

Confirmation of biochemical function

Although we do not know possible substrates for the putative methyltransferase MJ0882, we attempted to assay for its interactions with AdoMet. We incubated ^3H -labeled AdoMet with a 10-fold molar excess to purified MJ0882 protein and applied the mixture to a sephadex G-50 gel filtration column. In the absence of MJ0882, AdoMet elutes as a discrete peak after more than one column volume (Figure 4). In the presence of MJ0882, two peaks of ^3H -AdoMet were observed (Figure 4). The first peak eluted with the void volume and coincided exactly with the peak of MJ0882 elution, deduced by SDS-PAGE of the column fractions (data not shown). We observed a shoul-

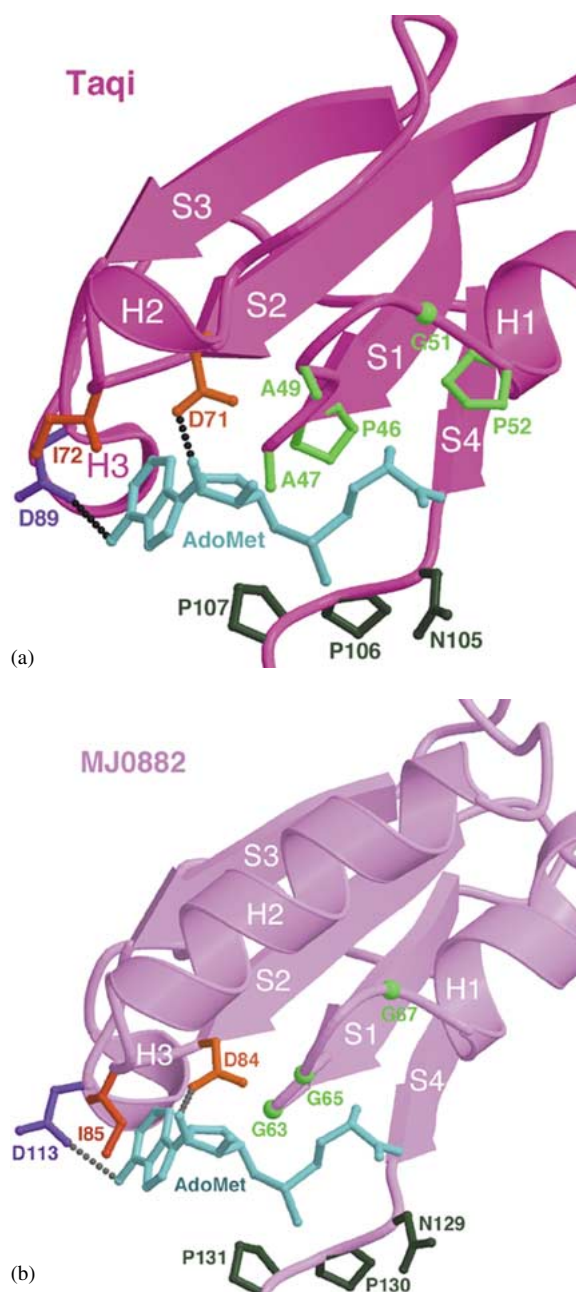


Figure 3. Comparison of AdoMet binding pockets of TaqI DNA methyltransferase and MJ0882 [prepared with the program MOLSCRIPT[23]]. (a) TaqI AdoMet binding pocket. AdoMet contacting side chains are colored and labeled. AdoMet is drawn in stick and ball model. Black dashed lines indicate hydrogen bonds. (b) Model for MJ0882 AdoMet binding pocket. AdoMet was modeled into the MJ0882 through structural alignment with TaqI. Possible AdoMet contacting side chains are colored and labeled. Black dashed lines indicate possible hydrogen bonds.

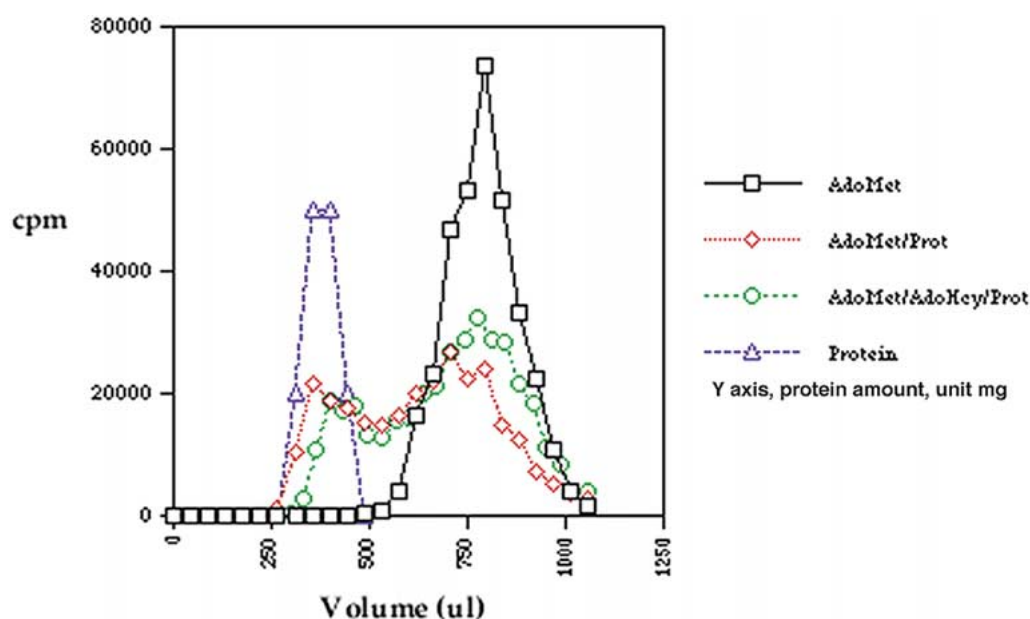


Figure 4. Analysis of MJ0882 binding to AdoMet and AdoHcy by gel filtration chromatography. MJ0882 was incubated with ^3H -labeled AdoMet and chromatographed through Sephadex G-50 as described in Materials and methods. The profile of the [^3H]AdoMet elution was determined by counting aliquots of each fraction. The profile of AdoMet is colored in black in the absence of MJ0882 and in red in the presence of MJ0882. The blue trace represents the elution profile of the MJ0882 protein as determined by SDS PAGE, protein concentration and volume. The effect of co-incubation of the AdoMet with AdoHcy at a 1 : 100 ratio is colored green.

der linking of this peak with the second peak whose migration was equivalent to that of unbound AdoMet. These data suggest a weak interaction between MJ0882 and AdoMet. The presence of considerable tailing of AdoMet between the two peaks despite the presence of cold AdoMet in the elution buffer suggests a moderate off rate in AdoMet binding to MJ0882. In addition, cold S-adenosylhomocysteine (AdoHcy), the de-methylated cofactor, is able to compete for the binding of AdoMet with MJ0882, at a concentration (100 : 1) relative to that of AdoMet (Figure 4). NAD does not compete with AdoMet's binding to MJ0882 (data not shown), which suggests that MJ0882 belongs to the methyltransferase family but not to the NAD dehydrogenase family.

The observed weak binding of AdoMet to MJ0882 is possibly due to the lack of a specific substrate. The protein loops important for AdoMet binding in both TaqI and HhaI DNA methyltransferases are unstable without their respective DNA substrates [15]. Since MJ0882 shares sequence similarity with that of putative RNA methyltransferase MTH724 (Figure 2b), RNA could be a potential substrate for MJ0882. If MJ0882 were a DNA or RNA methyltransferase, it would require an additional protein or domain to fa-

cilitate DNA or RNA binding, similar to those of the HhaI and TaqI methyltransferases. It is conceivable that such a DNA- or RNA-binding protein could be supplied in trans in the MJ organism.

In conclusion, we have presented an example of structure-based functional assignment of a hypothetical protein. The protein structure and preliminary biochemical data suggests that MJ0882 is a methyltransferase that binds to AdoMet, the common cofactor for methyltransferases. We are limited in our biochemical analysis by the absence of a known substrate.

Acknowledgements

This work was supported by grants from NIH (P50 GM62412) and DOE (The Director, Office of Science, Office of Biological and Environmental Research under U.S. Department of Energy Contract No. DE-AC03-76SF00098). We thank Jaru Jancarik for her selfless guidance in crystallization; David King for his help with mass spectrometry; Xiao-dong Cheng at Emory University for stimulating discussions and critical reading of the manuscript. The help

of the staff at the Advanced Light Source Beamline 5.0.2 is greatly appreciated.

References

- Schluckebier, G., O'Gara, M., Saenger, W. and Cheng, X. (1995) *J. Mol. Biol.* **247**, 16–20.
- Cheng, X., Kumar, S., Posfai, J., Pflugrath, J.W. and Roberts, R.J. (1993) *Cell* **74**, 299–307.
- Klimasauskas, S., Kumar, S., Roberts, R.J. and Cheng, X. (1994) *Cell* **76**, 357–369.
- Labahn, J. *et al.* (1994) *Proc. Natl. Acad. Sci. USA* **91**, 10957–10961.
- Reinisch, K.M., Chen, L., Verdine, G.L. and Lipscomb, W.N. (1995) *Cell* **82**, 143–153.
- Gong, W., O'Gara, M., Blumenthal, R.M. and Cheng, X. (1997) *Nucleic Acids Res* **25**, 2702–2715.
- Vidgren, J., Svensson, L.A. and Liljas, A. (1994) *Nature* **368**, 354–358.
- Fu, Z. *et al.* (1996) *Biochemistry* **35**, 11985–11993.
- Hodel, A.E., Gershon, P.D., Shi, X. and Quijcho, F.A. (1996) *Cell* **85**, 247–256.
- Bussiere, D.E. *et al.* (1998) *Biochemistry* **37**, 7103–7112.
- Djordjevic, S. and Stock, A.M. (1997) *Structure* **5**, 545–558.
- Aravind, L. and Koonin, E.V. (2001) *Trends Biochem. Sci.* **26**, 215–217.
- Holm, L. and Sander, C. (1993) *J. Mol. Biol.* **233**, 123–138.
- Hendrickson, W.A. (1991) *Science* **254**, 51–58.
- Cheng, X. (1995) *Annu. Rev. Biophys. Biomol. Struct.* **24**, 293–318.
- Kim, R. *et al.* (1998) *Biotech. Lett* **20**, 207–210.
- Jancarik, J., Scott, W.G., Milligan, D.L., Koshland, D.E., Jr. and Kim, S.H. (1991) *J. Mol. Biol.* **221**, 31–34.
- Otwinowski, Z. and Minor, W. (1997) *Methods Enzymol.* **276**, 307–326.
- Terwilliger, T.C. (1997) *Methods Enzymol.* **276**, 530.
- Cowtan, K. and Main, P. (1998) *Acta Crystallogr. D Biol. Crystallogr.* **54**, 487–493.
- Jones, T.A., Zou, J.-Y., Cowan, S.W. and Kjeldgaard, M. (1991) *Acta Crystallogr.* **A47**, 110–119.
- Brunger, A.T. *et al.* (1998) *Acta Crystallogr.* **D54**, 905–921.
- Kraulis, P.J. (1991) *J. Appl. Crystallogr.* **24**, 946–950.